Language Research Bulletin, *34*, ICU, Tokyo

Comparing the Effect of Two Extensive Reading Treatments on Receptive Vocabulary Knowledge

Kimberly Klassen
English for Liberal Arts
International Christian University

Michael Green
Seikei Institute for International Studies
Seikei University

Abstract

The aim of this quasi-experimental study is to investigate the effect of extensive reading (ER) on Japanese L1 learners' receptive knowledge of high frequency English vocabulary. This study adds to existing research into the impact of ER on lexical acquisition by comparing two groups who did ER for different lengths of time, and utilising a validated testing tool as part of the assessment. The vocabulary knowledge of participants (N = 78) was pre- and post-tested using the Vocabulary Levels Test. Group 1 did ER for ten weeks and studied abroad for six weeks; Group 2 did ER for 30 weeks and did not study abroad. A Mann-Whitney test found no statistically significant difference between the groups in terms of vocabulary gains. A small effect was found for the number of words read and vocabulary gains for Group 2. These results suggest there is no difference in high frequency vocabulary gains between the students who did extensive reading over a longer period, and those who did ER for a shorter period and studied abroad. Another important finding was that the learning gains were small in comparison to other published ER studies. Possible explanations for the small gains, and limitations, are discussed.

Extensive Reading (ER) is "reading in which learners read large quantities of material that are within their linguistic competence" (Grabe & Stoller, 2002, p. 259). In contrast, intensive reading involves assisted, detailed reading of texts which are more difficult due to subject matter, discourse markers, or vocabulary and grammar (Nation, 2008). The benefits of extensive reading on second language (L2) proficiency have long been strongly argued and researched (e.g. Davis, 1995; Day & Bamford, 1998; Grabe, 2009; Horst, 2009; Nation, 2009). For example, positive effects have been seen on learners' grammatical and lexical knowledge, as well as reading speed, reading comprehension and attitudes towards reading (see Nakanishi (2015) for a valuable meta-analysis of experimental and quasi-experimental ER studies). Indeed, ER has come to be seen as essential to language programs, and is viewed as a "crucial adjunct to classroom teaching" (Davis, 1995, p. 329). As Horst (2009) argues, in the absence of extensive reading, "key aspects of linguistic development cannot readily be acquired" (p. 41).

The aim of the study described here is to compare two treatments of ER on receptive knowledge of high frequency English vocabulary: treatment Group 1 received 10 weeks of ER, followed by six weeks of study abroad in an English speaking country; treatment Group 2 received 30 weeks of ER and did not study abroad. The study focused on high frequency

vocabulary (i.e. the first 3,000 words) because these words are very important for learners to acquire; the first 3,000 words constitute approximately 90% of the vocabulary in any given text, and can be considered basic English vocabulary (Schmitt & Schmitt, 2014). Additionally, the first 3,000 words are used by lexicographers of learner dictionaries for 'defining vocabulary' (Schmitt & Schmitt, 2014), that is, the vocabulary used in the definitions of key words. Therefore, it is crucial for learners to gain a solid understanding of this basic vocabulary to comprehend more difficult academic texts and to make effective use of learner materials. Graded readers (i.e. readers in which the vocabulary has been controlled for) typically utilise lexis up to the 3,000 level (Nation, 2018). For this reason, graded readers are seen as a valuable learning resource for high frequency vocabulary. Thus, there is the potential for learners who are engaged in ER programs to make gains in their knowledge of high frequency vocabulary.

Literature Review

Research findings on ER are being published at an ever increasing rate, and although reading speed and reading comprehension constitute the most common areas of interest, the beneficial effect of ER on vocabulary breadth and depth also commands attention. For example, of the 34 ER studies reviewed in Nakanishi (2015), 35% had vocabulary acquisition as the main area of research interest. Explicit vocabulary learning can be augmented by the adoption of an ER or reading fluency component, which Nation (2007) suggests should account for about one quarter of any L2 classroom curriculum. ER increases the exposure to intentionally learned lexis and enhances knowledge of the contextual usage of the target language as "the vocabulary load is both fine-tuned for the learner's level, and systematically recycled" (Schmitt, 2008, p. 349). This recycling of language can occur both within the graded reader text itself and the classroom context. In terms of incidental vocabulary acquisition, Coady (1997) asserts that the massive amount of exposure resulting from sustained reading-for-meaning leads to greater lexical growth than any classroom-based explicit instruction could ever achieve (p. 225).

The notion that ER is a vital component of vocabulary acquisition is backed by an accumulating amount of empirical data. Al-Nujaidi (2004) found that college students in a three-month ER treatment group gained 233 words at the 2000 level of the Vocabulary Levels Test (VLT) (Schmitt, Schmitt, & Clapham, 2001) and 283 words in the 3000 section. In contrast, the control group, which had no extra exposure to English texts, gained just 15 words and 48 words respectively, over the same three month period. Al-Homoud and Schmitt (2009) compared an ER class with a more traditional class involving intensive reading and vocabulary exercises over a 10 week period. The 70 participants were given a test battery pre- and post-treatment that consisted of the VLT, TOEFL and PET reading comprehension tests and a reading speed test. A questionnaire was also administered at the end of the treatment to investigate the participants' perceptions of their learning under the two reading approaches. The authors found that both groups made statistically significant gains at the 2000, 3000 and Academic Vocabulary levels of the VLT, though there was no statistically significant difference in the amount of gains between the two groups. The authors noted that the gains in academic vocabulary size made by the ER group is somewhat surprising, as graded readers typically have less academic lexis, especially at the lower levels used in their study. Where the two groups did diverge, however, was in their attitude towards the reading approach used; the questionnaire results showed that the participants in the ER group clearly had more positive attitudes toward their learning experience and gave a stronger indication of

learning than the participants in the intensive reading group (Al-Homoud & Schmitt, 2009, p. 398).

Additional robust evidence can be found in Lee (2007), which differs from the majority of ER investigations in two ways: it used a large control group (n = 139) and one experimental group (n = 67), and secondly, unlike most studies which are short-term (i.e. less than three months), involved a 12 month treatment. Vocabulary knowledge was pre-tested using the VLT, with no statistically significant differences among the groups. Over a complete academic year, the participants in the control group had text-book oriented instruction with explicit practice of intensive reading skills, presentations based on issues related to the assigned readings, plus regular quizzes and examinations. Meanwhile, the treatment group did self-selected reading of graded readers and completed reading logs recording time spent on reading, the number of pages read and reflections on the books' content. Post-testing using the same instrument (VLT) revealed that the experimental group outperformed the control group across all levels of the test and the differences in gains were statistically significant at all sections of the test apart from the academic vocabulary. Extrapolating from the reported mean gains suggests that participants in the ER group learned approximately 475 words over the academic year, compared to 240 words in the control group. In addition to ER participants reading faster, reading more and comprehending better, what Nuttall (2005) calls the "virtuous circle of the good reader" (p. 127), research also indicates that ER facilitates vocabulary acquisition. Nakanishi's (2015) report of a large effect for vocabulary *(d* = 1.25) needs careful interpretation however, as the results are from pre-post contrast analysis of a relatively small number of studies with no control groups.

The study described here addresses these two research questions:

Research Question 1 (RQ1): What is the difference in terms of receptive vocabulary gains of high frequency words in two treatments: extensive reading over ten weeks plus study abroad for six weeks, and extensive reading over 30 weeks without a study abroad period?

Research Question 2 (RQ2): What is the relationship between the total number of words read in the 30-week ER course and gains of high frequency vocabulary?

Method

Participants

Participants were first-year Japanese learners of English at a private bilingual university. There were a total of 124 students in the low-intermediate proficiency stream (TOEFL 350-450) in the academic year of 2018. However, data from 46 participants was removed either because they had missed a pre- or post-test, or their test scores were not collected by a teacher. This left data from 78 participants for the study. The participants were enrolled in a total of 22 credit hours of English courses over the academic year (Reading Content and Analysis; Academic Reading and Writing; and Academic Skills).

Testing Instruments

To measure participants' vocabulary knowledge, the Vocabulary Levels Test (Schmitt et al., 2001) was used. The VLT has been used in numerous studies, including four studies from Nakanishi's (2015) ER meta-analysis, in which he encourages researchers to use the same tests so that valid comparisons can be made between results. Al-Homoud and Schmitt (2009) further note that the VLT is considered a standard test for testing vocabulary knowledge (p. 389, 390). In the present study, Version 2 of the VLT was used at the 2,000

(2K) and 3,000 (3K) levels. The Academic Word List was not tested, as the focus of this study is on the effect of ER on high frequency vocabulary (i.e. the first 3,000 words). Schmitt et al. (2001) report reliability figures of .922 for Version 2 of the VLT.

Procedure

Pre-test vocabulary scores were collected in the first ER class of the spring term in early April: all participants were given the VLT 2K and 3K levels to complete. They were not allowed to use dictionaries and were told not to guess if there was an unknown word on the test. They were given 15 minutes to complete the two levels, though most participants finished in ten minutes. Tests were graded by classroom teachers (using an answer key) and recorded.

The participants were then introduced to extensive reading in the spring term. The ER course participants met once a week for 70 minutes. Students were given training in how to choose appropriate books for their proficiency level, and were given some class time to do Sustained Silent Reading every week. They also did a timed reading course in order to improve their reading speed. Participants kept track of the number of words they read in ER classes using a free online tool designed for this purpose (www.mreader.org). The website gauges reading comprehension with short quizzes: if a student gets more than 60% on a quiz, that book is counted toward their word goal. The word count goal for the spring term was 70,000 words over a ten-week period. Participants were also asked to make word cards for ten self-selected, new (i.e. previously unknown) vocabulary items from their graded readers every week, for a total of 100 words over the term.

Then, of the 78 participants, 40 studied abroad (Group 1) over the summer break in a university-run ESL program in an English speaking country, for an average of six weeks. They returned for the autumn and winter terms but did not do any further ER. The other 38 students (Group 2) did not study English abroad during the summer, and continued the ER course for the autumn and winter terms (of 10 weeks each). The ER course met once a week and participants continued to self-select graded readers and make ten new word cards to learn each week. No class time was given for silent reading; instead, class time was used for small group discussions of books they read, word card review and reading fluency practice. The word count goal for the autumn term was 100,000; the goal for the winter term was 130,000 words. Thus, in total, Group 2 participants were encouraged to read 300,000 words over the academic year while Group 1 participants were encouraged to read 70,000 words over the spring term only.

Post-test vocabulary scores were collected in the last week of the winter term in late February and the same procedure for administering the test and recording scores was followed as for the pre-test score collection.

Analysis

To investigate the effect of extensive reading on knowledge of high frequency vocabulary, an independent samples t-test run was run to compare Group 1 and Group 2 gains at the 2K and 3K levels. A correlation matrix was also generated to investigate the relationship between total words read in the ER course and vocabulary gains at the end of the academic year.

Results

In order to address RQ1 (What is the difference in terms of receptive vocabulary gains of high frequency words in two treatments: extensive reading over ten weeks plus study

abroad for six weeks, and extensive reading over 30 weeks without a study abroad period?), descriptive statistics were compiled for the 2K and 3K levels separately, by group. Table 1 shows the descriptive statistics for the 2,000 (2K) level of the Vocabulary Levels Test. The total score of the 2K test is 30.

Table 1
Descriptive statistics for 2K level in April (pre-test) and February (post-test) by groups

Group		April		February		Gains				
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	Range	Min.	Max.
1[1]	40	27.02	2.13	27.73	1.81	1.20	1.48	6.00	0.00	6.00
2[2]	38	26.47	2.87	27.68	1.92	1.47	1.94	7.00	0.00	7.00

[1] *ER for ten weeks and study abroad for six weeks*
[2] *ER for 30 weeks and no study abroad*

Initial inspection of the data indicated that the assumptions of parametric testing were not met. Data was not normally distributed, as indicated by the presence of outliers, the Shapiro-Wilk test and skewness values over 1. For this reason, a non-parametric test was run. A Mann-Whitney test indicated that there was no statistically significant difference in 2K vocabulary gains from April to February for Group 1 (*Mdn* = 1) and Group 2 (*Mdn* = 1), $U = 791.50$, $p = 0.74$, $r_b = 0.04$. Extrapolating from the mean gains, Group 1 learned 40 words from the 2K level, while Group 2 gained on average 49 words from this band over the academic year.

Table 2 shows the descriptive statistics for the 3,000 (3K) level of the VLT. The total score of the 3K test is 30.

Table 2
Descriptive statistics for 3K level in April (pre-test) and February (post-test) by groups

Group		April		February		Gains				
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	Range	Min.	Max.
1[1]	40	22.52	4.80	23.73	3.62	2.05	2.86	12.00	0.00	12.00
2[2]	38	21.11	5.57	22.68	4.26	2.53	3.27	13.00	0.00	13.00

[1] *ER for ten weeks and study abroad for six weeks*
[2] *ER for 30 weeks and no study abroad*

Initial inspection of the data indicated that assumptions of parametric testing were not met. Boxplots suggested that variances were unequal, and the data was not normally distributed, as indicated by the presence of outliers, the Shapiro-Wilk test and skewness

values over 1. A Mann-Whitney test found no statistically significant difference in 3K vocabulary gains from April to February for Group 1 (*Mdn* = 1) and Group 2 (*Mdn* = 1), U = 807.50, p = 0.63, r_b = 0.06. Extrapolating from the mean gains, Group 1 learned approximately 68 words on average from this band over the academic year; Group 2 learned on average 84 words from the 3K level.

For RQ2, in order to investigate the relationship between ER and knowledge of high frequency vocabulary, descriptive statistics were compiled to determine how much extensive reading was done by Group 2 over the academic year in terms of word counts, as seen in Table 3.

Table 3
Descriptive statistics for ER word counts from Group 2 for AY 2018

Group	*n*	*M*	*SD*	Range	Min.	Max.
2[1]	38	289,200	73,770	428,300	102,400	530,700

[1] *ER for 30 weeks and no study abroad*

Recall that the word count goal for Group 2 was 300,000 for the entire academic year. The average word count in that group was just under that goal (289,200). Again, there was much variation in the group, with the highest participant achieving over half a million words read.

A Pearson's r correlation (one-tailed) between the number of words read and gains at the 2K level for Group 2 was statistically significant with a small effect size (95% CI: .12, 1.00, p = .008, r = .39, R^2 = 0.15). The correlation between the number of words read and gains at the 3K level for Group 2 was also statistically significant, with a small effect size (95% CI: .12, 1.00, p = .009, r = .38, R^2 = 0.14).

In sum, the answer to RQ2 is that there is a statistically significant relationship between the number of words read and vocabulary gains, albeit with a small effect size, but this, of course, does not imply a causal relationship.

Discussion

The results show that Group 2 who did ER for the whole academic year did not make statistically significant gains in their knowledge of high frequency vocabulary. The participants in Group 2 had an average net gain of approximately 133 words in their receptive knowledge of high-frequency words (2K and 3K combined). In comparison, a similar result was found for Group 1 who did ER for ten weeks and studied abroad for six weeks. Group 1 gained on average 108 high frequency words (2K and 3K combined). The difference in gains between the two treatment groups was not statistically significant. While the correlation between the number of words read and vocabulary gains for the Group 2 was statistically significant, the effect sizes were small at both the 2K and 3K levels.

Looking only at the 2K level, it is perhaps unsurprising that there were few learning gains. From the April (pre-test) scores, we can see that average scores for both groups was 27 out of 30. When interpreting the results from the VLT, it is generally assumed that a score of 90% or more (i.e. 27 or greater) suggests that the level is 'known' (Nation, 2008, p. 143).

Thus, on average, participants in both groups knew the 2K level, and therefore, few gains would be expected. Lee (2007) found a similar ceiling effect for the 2K level.

Turning to the gains for the 3K level, it is somewhat disconcerting that the increases were not greater for this band over the 30-week period of academic study. Recall that Group 2 learned on average 84 words from this level, while Group 1 learned 68 words on average. As discussed above, knowledge of high frequency vocabulary is crucial for L2 learners: 90% of lexical items in any text they encounter will be comprised of high frequency vocabulary, and knowledge of this basic vocabulary is important for effective use of learner materials, such as dictionaries. The mean post-test scores (i.e. February; see Table 2) indicate that this level of vocabulary is not well known: for Group 2, we can extrapolate that about 250 words of the 1,000 at this level are unknown, on average. For Group 1, we can estimate that about 210 words at the 3K level are not known. Thus, there remains a significant portion of essential vocabulary to be learned.

Looking at combined gains from both the 2K and 3K levels, the vocabulary gains reported here are small in comparison to the studies reviewed above. For example, in Al-Nujaidi's (2004) study, combined gains (2K and 3K) were 516 words over three months. In Lee (2007), combined gains for the ER group were 475 words over a 12-month period, compared to 240 words in the control groups. Thus, the gains reported here for Group 2 (133 words) and for Group 1 (108 words) seem comparatively minor, especially when taking into account the length of study.

One possible explanation for the results can be found in Nation (2008), where he points out that because vocabulary learning from ER is largely incidental, such gains can be fragile (p. 70). For this reason, he stresses the importance of setting the quantity of input sufficiently high to ensure repeated exposure to high frequency items: close to 500,000 running words over a year. In that regard, the word count goals of the ER course may have been set too low at 300,000 over the academic year. In a similar vein, Nagy (1997) suggests that reading a million words over an academic year allows for a potential gain of 1,000 words. In the results reported here, the participants in Group 2 read close to 290,000 words on average. Therefore, we might expect their combined gains to be closer to 300 words, not the 133 words as was found here. Gains of 300 words in one academic year would also be more in line with results found in other ER studies reviewed above.

It is also worth noting that participants in both groups may have learned vocabulary that was not captured by the VLT (i.e. mid and low frequency vocabulary beyond the 3K level). While the vocabulary in graded readers is mostly comprised of the first 3,000 words, publishers of graded readers tend to categorise their readers by number of headwords (Schmitt & Schmitt, 2014). For that reason, mid and low frequency vocabulary items can be found in graded readers. As previously mentioned, the participants in the ER course self-selected vocabulary to learn from their graded readers. Thus, they may have learned mid or low frequency vocabulary. Future research could investigate this possibility by including the AWL (academic word list) and 5K levels of the VLT, which were not tested here because the effect of ER on high frequency vocabulary was the study's focus.

Moreover, it is important to note that individual learner differences inevitably get lost in statistical analyses. Thus, while the average vocabulary gain seems low for Group 2, there were participants who made much larger vocabulary advancements. For example, the participant who read over half a million words (see Table 3) made extrapolated gains of 566 words over the academic year, which is much more in line with the reported gains in other ER studies, as well as the potential learning estimates from Nagy (1997) and Nation (2001). Furthermore, while the participants in Group 2 read close to 300,000 words on average over the year, some participants read much less (see Table 3). For example, the participant who read the least (just over 100,000 words) made zero gains at both the 2K and 3K levels.

There are several limitations to the study, a few of which are mentioned here. First, and most importantly, the sample size for the study was small with under 50 participants in each group. The absence of a statistically significant finding could well be the result of the study having low statistical power. The second limitation is also related to the sample population: one teacher in the program chose not to collect vocabulary post-test scores for some participants, thus there was a loss of data which could have impacted the results. Third, there were several uncontrolled variables in the study. For example, some of the Group 1 participants may have continued ER in their own time after having been introduced to it in the spring term. Also, while the ER participants were asked to make 100 word cards per term, it is unknown how much time was spent reviewing these word cards. It should be noted that because of the word card task, the vocabulary learning was not entirely incidental; rather, participants were encouraged to deliberately learn the new words they had self-selected from their graded readers. Future research could control for such a variable by ensuring that in each class meeting, a certain amount of time was devoted to word card review.

Conclusion

Unlike other studies investigating the effect of ER on high frequency vocabulary gains, the study reported here did not find a statistically significant difference between gains made by the two treatment groups. Furthermore, the gains made by each group were relatively small in comparison to results in other ER studies (e.g. Al-Homoud & Schmitt, 2000; Al-Nujaidi, 2004; Lee, 2007). A few possible reasons for these disparate results were mentioned, including problems related to the sampling. For that reason, it may be valuable to repeat the experiment with a complete representative sample of participants in the low-intermediate stream.

Also, raising the word count goals in the ER course to half a million words over the year is worthy of consideration. Certainly, there is much potential for vocabulary learning through ER. Nation (2001) calculates that if a learner reads a million words in an academic year with 96-98% words known (i.e. extensive reading), they will be exposed to 20,000 to 40,000 new words. He points out that if the learner learned just one in ten of these new words, they could learn between 2,000 and 4,000 a year. That might be an ambitious goal, but the calculation does demonstrate the vast potential for vocabulary learning through ER. A million words (i.e. approximately 65 graded readers) in an academic year is entirely feasible: Nation (2001) computes that a learner could reach this goal even by reading at a relatively slow speed of 100 wpm, for 45 minutes a day, 222 days of the year. Previous research has demonstrated that learners at this proficiency level can achieve reading rates of 215 wpm on average after completing a speed reading course (Klassen & Allan, 2018). Thus, raising the word count goal in the ER course to even half a million words is very achievable: this would require reading at 200 wpm for 12 minutes every day, for 210 days of the academic year. Given the possible vocabulary learning gains, it bears revisiting this goal of the ER course, while taking into account other academic demands in the program.

References

Al-Homoud, F., & Schmitt, N. (2009). Extensive reading in a challenging environment: a comparison of extensive and intensive reading approaches in Saudi Arabia. *Language Teaching Research, 13*(4), 383-401. doi:10.1177/1362168809341508

Al-Nujaidi, A. (2004). *The impact of extensive reading on EFL vocabulary development*. Paper presented at the TESOL Arabia Conference, Dubai.

Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition*. Cambridge: Cambridge University Press.

Davis, C. (1995). Extensive reading: An expensive extravagance? *ELT Journal, 49*, 329-336. doi:10.1093/elt/49.4.329

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge: Cambridge University Press.

Grabe, W., & Stoller, F. L. (2002). *Teaching and researching reading*. Harlow, England: Pearson Education.

Horst, M. (2009). Developing definitional vocabulary knowledge and lexical access speed through extensive reading. In Z. Han & N. J. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 40-63). Ann Arbor: University of Michigan.

Klassen, K. & Allan, T. (2018). Evaluating an extensive reading course. *Language Research Bulletin, 33,* 22-33.

Lee, S. Y. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal, 38*, 150-170. doi:10.1177/0033688207079730

Nagy, W. E. (1997). On the role of context in first and second language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly, 49*(1), 6-37. doi:10.1002/tesq.157

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*, 2-13. doi:10.2167/illt039.0

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. USA: Heinle Cengage Learning.

Nation, I. S. P. (2009). *Teaching ESL/EFL Reading and Writing*. New York: Routledge.

Nation, I. S. P. (2018). Reading a whole book to learn vocabulary. *ITL - International Journal of Applied Linguistics, 169*(1), 30-43. doi:10.1075/itl.00005.nat

Nuttall, C. (2005). *Teaching reading skills in a foreign language*. Oxford: Macmillan Education.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*, 329-363. doi:10.1177/1362168808089921

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484-503. doi:10.1017/S0261444812000018

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88. doi:10.1177/026553220101800103